




# Classifier for centrality determination with Zero Degree Calorimeter at the Cooling-Storage-Ring External-target Experiment\*

Biao Zhang <sup>1</sup>, Li-Ke Liu <sup>1</sup>, Hua Pei,<sup>1</sup> Shu-Su Shi,<sup>1</sup> Nu Xu,<sup>1</sup> and Ya-Ping Wang <sup>1,†</sup>

<sup>1</sup>Key Laboratory of Quark & Lepton Physics (MOE) and Institute of Particle Physics,  
Central China Normal University, Wuhan 430079, China

The zero-degree calorimeter (ZDC) plays a crucial role toward determining the centrality in the Cooling-Storage-Ring External-target Experiment (CEE) at the Heavy Ion Research Facility in Lanzhou (HIRFL). A boosted decision tree (BDT) multi-classification algorithm was employed to classify the centrality of the collision events based on the raw features from ZDC such as the number of fired channels and deposited energy. The data from simulated  $^{238}\text{U} + ^{238}\text{U}$  collisions at 500 MeV/u, generated by the IQMD event generator and subsequently modeled using the GEANT4 package, were employed to train and test the BDT model. The results showed the high accuracy of the multiclassification model adopted in ZDC for centrality determination, which is robust against variations in different factors of detector geometry and response. This study demonstrates the good performance of CEE-ZDC in determining the centrality in nucleus-nucleus collisions.

Keywords: ZDC, Boosted Decision Trees, Multi-classification, IQMD, Centrality determination

## I. INTRODUCTION

The primary objective of conducting heavy-ion collisions at different beam energies is to investigate strong interaction matter and understand the QCD phase diagram. The phase diagram provides information on the phase transition and critical point of a strongly interacting system, where hadron gases exist at lower temperatures and low baryon densities; at higher temperatures or densities, the hadronic boundary disappears, and confined quarks move freely throughout the system [1]. The Beam Energy Scan program of RHIC-STAR aims to approach the possible critical point from the high-energy side. However, it is essential to study the phase diagram of the hadron phase and approach the critical point from the low-energy side [2–4]. The Cooling-Storage-Ring External-target Experiment (CEE) at the Heavy Ion Research Facility in Lanzhou (HIRFL), with its advanced spectrometer, provides significant opportunities for studying phase diagrams at extremely high net baryon density levels with energies of several hundred AMeV [5].

The zero-degree calorimeter (ZDC), one of the subdetectors of CEE in the forward rapidity region, is designed to accurately determine the centrality and reaction plane of collision events [6]. Collision events are typically classified into centrality classes representing certain fractions of the total reaction cross-section corresponding to specific intervals of the impact parameter  $b$  [7]. The impact parameter  $b$  is essential for understanding the initial overlap region of the colliding nuclei in heavy-ion collisions; it represents the distance between the nuclei centers in the plane transverse to the beam axis and determines the size and shape of the resulting medium. However, the impact parameter  $b$  is not di-

rectly measurable in the experiments. To estimate centrality experimentally, raw observables that scale monotonically with impact parameter can be used for classification according to centrality, for example, the reconstructed tracks with central barrel tracking detectors or the deposited energy in the forward calorimeters. Accurate centrality determination is a baseline for many physical analyses in heavy-ion collision experiments [8], particularly when searching for observables sensitive to a possible phase transition or critical point through analysis of fluctuations and correlations.

In recent years, Machine Learning (ML) methods have gained significant attention for determining the centrality of heavy ion collisions [8, 9]. Previous studies treated centrality determination as a regression problem on impact parameters and utilized combined information from central tracking systems and forward calorimeters to train ML models. However, to avoid autocorrelation in the physics analysis, this study adopts a machine-learning approach that utilizes raw experimental features from a forward calorimeter to determine centrality. We report the application of a multi-classification ML algorithm based on Boosted Decision Trees (BDT) as a centrality classifier using only ZDC in  $^{238}\text{U} + ^{238}\text{U}$  collisions at 500 MeV/u at the CEE. The ML inputs were generated using the Isospin dependent Quantum Molecular Dynamics (IQMD) generator [10]. In addition, we present the efficiency and purity measures related to the centrality determination performance of the ZDC with a model application.

## II. CEE-ZDC

The CEE, which utilizes fixed-target-mode heavy-ion collisions, is the first large-scale experimental nuclear device operating in the GeV energy region in China. It is equipped with a set of sub-detectors, as shown in Fig. 1(a). The detector system comprises a beam monitor, T0 detector [5], time projection chamber (TPC) [11], inner time-of-flight (iTOF) detector [12], large superconducting dipole magnet, multiwire drift chamber (MWDC) [13], external time-of-flight (eTOF) detector [14], and zero-degree calorimeter (ZDC) [6].

\* This work was supported in part by the National Nature Science Foundation of China (NSFC) under Grant No. 11927901 and 12175084, the National Key Research and Development Program of China under Grant No. 2020YFE0202002 and 2022YFA1604900, and the Fundamental Research Funds for the Central Universities (No. CCNU22QN005).

† Corresponding author, [wangyaping@mail.ccnu.edu.cn](mailto:wangyaping@mail.ccnu.edu.cn)

The purpose of the ZDC is to detect particle fragments in the forward rapidity region following semi-central and peripheral collisions, which provides vital information for the precise reconstruction of the centrality and reaction plane of collision events [6, 15]. The ZDC is centrally positioned at the end of the CEE, covering a pseudorapidity range of  $1.8 < \eta < 4.8$ . The ZDC utilizes a symmetrical and fan-shaped layout with eight radial and 24 angular sections and a maximum radius of 1 m. The detector comprises trapezoidal modules equipped with uniform plastic scintillators that are coupled with a light guide and connected to photomultiplier tubes (PMT) to convert scintillation light into charge signals. To obtain a comprehensive signal, each module provides two charge signals from the two dynodes of each PMT that are transmitted to two separate readout channels, resulting in 384 ( $24 \times 8 \times 2$ ) channels for the ZDC.

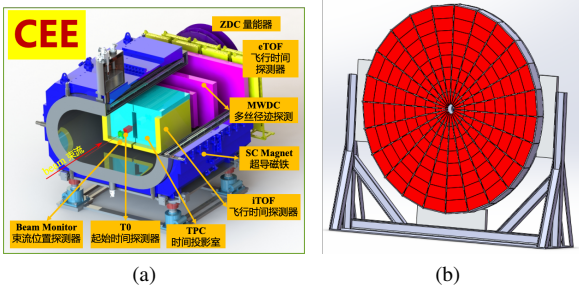


Fig. 1. (Color online) (a) CEE detector schematic layout. (b) ZDC detector layout.

### III. MODEL TRAINING WITH SIMULATED EVENT

The simulated data were generated by simulating  $^{238}\text{U} + ^{238}\text{U}$  collisions at 500 MeV/u using an IQMD generator [10]. The generated particles were then transported through the apparatus using the GEANT4 package [16]. Determining the centrality with only one forward rapidity detector, such as ZDC, is challenging even when employing ML algorithms. Previous ML-based studies on centrality determination relied on information from multiple subsystems within the detector, such as tracks reconstructed from central barrel detectors and deposited energy in forward calorimeters, revealing a strong correlation between the centrality class and observables. CEE-ZDC is a nontracking detector, and the number of spectator nucleons in a nucleus-nucleus collision is expected to be proportional to the deposited energy and number of fired channels in the ZDC. However, the presence of a beam hole at the center of ZDC and the limited detector acceptance result in a weak monotonic dependence between the impact parameters and observables, as illustrated in Fig. 2(a), which shows the number of fired channels and Fig. 2(b), which shows the energy deposited in ZDC.

Potential improvements in centrality determination can be achieved by utilizing data from ZDC-subrings in conjunction with the ZDC as an additional feature in the ML task. More-

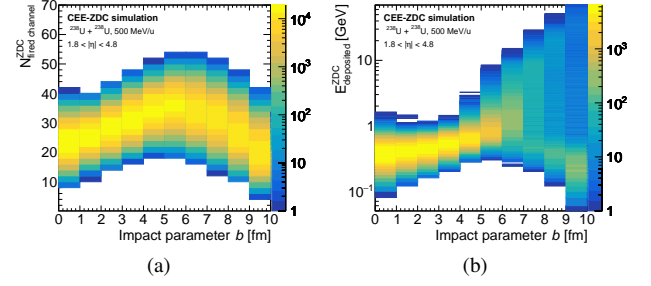


Fig. 2. (Color online) (a) The number of fired channels in ZDC as a function of impact parameter. (b) The deposited energy in ZDC as a function of impact parameter.

over, it may be advantageous to use the energy deposited in the ZDC ring-by-ring and the number of event-by-event fired channels and to exploit all inherent correlations between modules. Fig. 3(a) displays the probability distribution of the fired ZDC channels in the impact parameter range  $7 < b \leq 10$  fm as well as the probability distribution of the deposited energy of ZDC rings in the impact parameter range  $0 < b \leq 3$  fm shown in Fig. 3(b). The complex pattern and nontrivial decision boundary among the event centrality classes present an ideal opportunity for applying ML techniques.

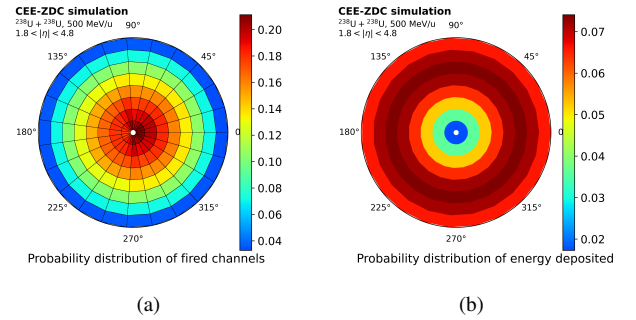


Fig. 3. (Color online) (a) Probability distribution of fired ZDC channels in impact parameter interval of  $7 < b \leq 10$  fm. (b) Probability distribution of deposited energy of ZDC rings in impact parameter interval of  $0 < b \leq 3$  fm.

Boosted Decision Trees (BDT), a family of popular supervised learning algorithms for classification and regression problems, are extensively used to analyze data in high-energy physics experiments. In this study, extreme gradient boosting (XGboost), a powerful BDTs based on the gradient boosting method, was adopted to solve multi-classification problems for centrality determination. The physical features used as the inputs for model training are the deposited energy in the full ZDC and ZDC substrings as well as the number of fired channels in ZDC. The simulated data were divided into three centrality classes based on the impact parameters listed in Table. 1. The samples were divided into training and test samples of equal size for each centrality class. A state-of-the-art machine learning hyperparameter optimization with Optuna

was adopted to speed up optimization time and achieve the best performance of the training models [17].

TABLE 1. The centrality classes with respect to the impact parameter  $b$  intervals

Centrality class	$b$ interval (fm)
Central	$0 < b \leq 3$
Semi-Central	$3 < b \leq 7$
Peripheral	$7 < b \leq 10$

#### IV. PERFORMANCE OF THE ML MODELS

The machine learning model was applied to both the training and test sets to visualize the distributions of the ML output scores and to check for consistency between the two sets. For classification with three centrality classes ( $p_i$ ), the model generates three scores representing the probability of belonging to each class considered. According to construction, the probabilities for the centrality classes sum to one ( $\sum_{i=1}^3 p_i = 1$ ). Fig. 4 illustrates the probability distributions of the central (a) and peripheral classes (b) for both the training and test sets. The probability distributions were close to unity for each probability distribution corresponding to the respective true class, whereas the other two distributions shifted toward zero. The probability density functions of the training and test samples for each centrality class agreed well, indicating that the model did not overfit.

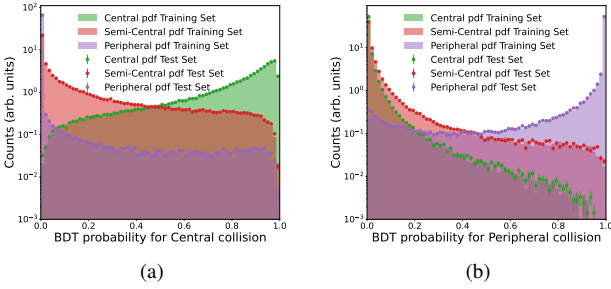


Fig. 4. (Color online) The probability distributions of belonging to the central class (a) and peripheral class (b) for both the training and test sets.

The Receiver Operating Characteristic (ROC) curve is commonly used to evaluate the performance of a classification model by plotting the true-positive rate against the false-positive rate for various threshold settings. The area under the ROC curve, known as ROC AUC, provides a global measure of the model performance, ranging from 0.5 (random classification) to 1 (perfect classification), independent of the threshold and class distribution [18]. However, for multi-class classification, the ROC curve cannot be directly defined, and the "One-vs-One" approach is used to compute the overall average of the individual ROC AUCs for each pair of classes. The ROC curves and ROC AUC values obtained for the test set are

shown in Fig. 5. The high final ROC AUC value of approximately 0.96 indicates that the BDT model is highly effective in determining centrality.

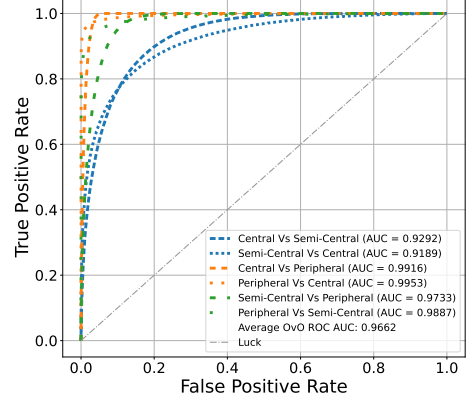


Fig. 5. (Color online) ROC curves and AUCs with respect to different "One-vs-One" cases are shown with the different line colors.

#### V. EFFICIENCY AND PURITY OF THE CENTRALITY CLASSIFICATION

The performance of the centrality classification model was evaluated by calculating its efficiency and purity based on ML output scores. Efficiency refers to the fraction of correctly classified events, whereas purity measures the fraction of events correctly classified for a particular centrality class out of all the events assigned to that class. The efficiency versus purity of the multiclassification models for each centrality class is shown in Fig. 6, where the red, green, and blue solid lines represent the central, semicentral, and peripheral classes, respectively. The peripheral class was the most effectively classified, and the central class was more challenging than the semi-central class in higher-efficiency regions. The values listed in Table. 2 indicate that even at very high purity levels, the efficiency of the peripheral class is not significantly compromised, and both the central and semi-central classes exhibit promising efficiency values at high purity. These results indicate that the ML-based event centrality determination utilized in ZDC is effective.

TABLE 2. Efficiency and purity values for different centrality classes

Efficiency \ Purity \ Class	Central	Semi-Central	Peripheral
Purity = 90%	67%	66%	97%
Purity = 95%	41%	47%	94%
Purity = 98%	11%	24%	93%

In addition, to evaluate the performance of the centrality determination with ZDC, the effects of several factors

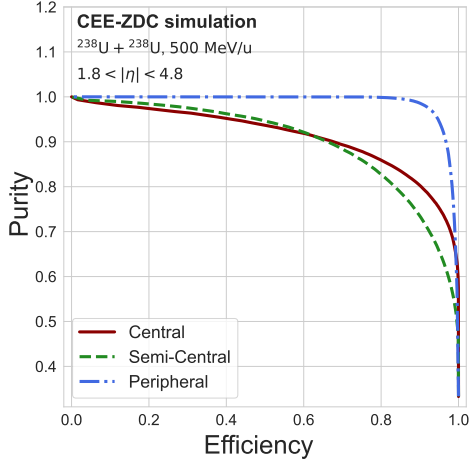


Fig. 6. (Color online) Efficiency versus purity of the multi-classification models for each centrality class. The red, green, and blue lines represent the central, semi-central, and peripheral classes, respectively.

related to the configuration of ZDC in the simulation data were systematically investigated. These factors include the thickness of the plastic scintillator of ZDC detector, hit efficiency, energy resolution, and heavy nuclei with or without de-excitation (tunable settings in IQMD). The ZDC plastic scintillator thickness was varied from 1 to 4 cm, and the hit efficiency was varied from 90% to 95%. The deposited energy was also smeared with different sigma values of Gaussian distributions. As illustrated in Fig. 7, the red, green, and blue lines indicate central, semi-central, and peripheral collisions, respectively. Changes in these factors are depicted by distinct line styles. The results indicated that the effects of these factors on the purity and efficiency of the centrality classification were minor. Among the tested factors, the ZDC detector thickness had the most significant impact, although its effect was relatively small. In conclusion, this study suggests that the multi-classification adopted in ZDC is robust against variations in these factors, indicating the potential for reliable and accurate classification of centrality using ZDC.

## VI. SUMMARY

This study aimed to determine the centrality class of nucleus-nucleus collisions at the CEE-ZDC detector using a multi-classification model based on the XGBoost classi-

fier. The ML model was trained and tested using simulation data from the IQMD event generator, and then modeled using the GEANT4 package. An additional study examined various factors associated with the geometry and response of the ZDC detector. The results indicated that the impact of these factors was minor, demonstrating the robustness of the XGBoost classifier in determining centrality. Future work may include improving the accuracy of centrality determination by incorporating regression tasks and exploring other machine-learning algorithms. This study indicates the good performance of CEE-ZDC for centrality determination in nucleus-nucleus collisions.

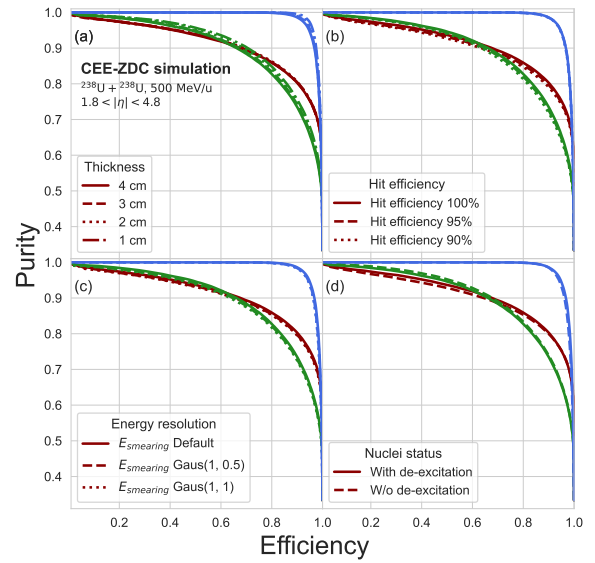


Fig. 7. (Color online) The effects of several factors on the efficiency and purity for the multi-classification models: (a) thickness of ZDC detector, (b) hit efficiency in ZDC (b), (c) energy resolution, (d) with or without de-excitation. The red, green, and blue colors of the lines represent central, semi-central, and peripheral collisions, respectively. The variation of the factors is shown with the different line styles.

## ACKNOWLEDGMENTS

We thank Prof. Li Ou and Zhigang Xiao for generating IQMD data and for fruitful discussions.

- [1] P.B. Munzinger and J. Stachel, The quest for the quark-gluon plasma. *Nature* **448**, 302–309 (2007). doi: [10.1038/nature06080](https://doi.org/10.1038/nature06080)
- [2] D. Oliinychenko, A. Sorensen, V. Koch et al., Sensitivity of Au+Au collisions to the symmetric nuclear matter equation of

- state at 2 – 5 nuclear saturation densities. [arXiv:2208.11996](https://arxiv.org/abs/2208.11996)
- [3] H.Z. Huang, F. Liu, X.F. Luo *et al.*, Collective excitation in high-energy nuclear collisions – in memory of Professor Lianshou Liu. *Symmetry* **15**, 499 (2023). doi: [10.3390/sym15020499](https://doi.org/10.3390/sym15020499)

- [4] X. Luo, S. Shi, N. Xu *et al.*, A study of the properties of the QCD phase diagram in high-energy nuclear collisions. *Particles* **3**, 278–307 (2020). doi: [10.3390/particles3020022](https://doi.org/10.3390/particles3020022)
- [5] D.D. Hu, J.M. Lu, J. Zhou *et al.*, Extensive beam test study of prototype MRPCs for the T0 detector at the CSR external-target experiment. *Eur. Phys. J. C* **80**, 282 (2020). doi: [10.1140/epjc/s10052-020-7804-2](https://doi.org/10.1140/epjc/s10052-020-7804-2)
- [6] S.H. Zhu, H.B. Yang, H. Pei *et al.*, Prototype design of read-out electronics for Zero Degree Calorimeter in the HIRFL-CSR external-target experiment. *JINST* **16**, P08014 (2021). doi: [10.1088/1748-0221/16/08/P08014](https://doi.org/10.1088/1748-0221/16/08/P08014)
- [7] A. Svetlichnyi, R. Nepeyvoda, I. Pshenichnov, Using Spectator Matter for Centrality Determination in Nucleus-Nucleus Collisions. *JINST* **4**, 227–235 (2021). doi: [10.3390/particles4020021](https://doi.org/10.3390/particles4020021)
- [8] A. Igor and K. Vladimir, Classifiers for centrality determination in proton-nucleus and nucleus-nucleus collisions. *EPJ Web Conf.* **137**, 11001 (2017). doi: [10.1051/epjconf/201713711001](https://doi.org/10.1051/epjconf/201713711001)
- [9] M. Omana Kuttan, J. Steinheimer, K. Zhou, A. Redelbach, H. Stoecker, A fast centrality-meter for heavy-ion collisions at the CBM experiment. *Phys. Lett. B* **811**, 135872 (2020). doi: [10.1016/j.physletb.2020.135872](https://doi.org/10.1016/j.physletb.2020.135872)
- [10] C. Hartnack, R.K. Puri, J. Aichelin *et al.*, Modeling the many body dynamics of heavy ion collisions: Present status and future perspective. *Eur. Phys. J. A* **1**, 151-169 (1998). doi: [10.1007/s100500050045](https://doi.org/10.1007/s100500050045)
- [11] W. Huang, F. Lu, H. Li *et al.*, Laser test of the prototype of CEE time projection chamber. *Nucl. Sci. Tech.* **29**, 41 (2018). doi: [10.1007/s41365-018-0382-4](https://doi.org/10.1007/s41365-018-0382-4)
- [12] X. Wang, D. Hu, M. Shao *et al.*, CEE inner TOF prototype design and preliminary test results. *JINST* **17**, P09023 (2022). doi: [10.1088/1748-0221/17/09/P09023](https://doi.org/10.1088/1748-0221/17/09/P09023)
- [13] L. Lyu, H. Yi, L. Duan *et al.*, Simulation and prototype testing of multi-wire drift chamber arrays for the CEE. *Nucl. Sci. Tech.* **31**, 11 (2020). doi: [10.1007/s41365-019-0716-x](https://doi.org/10.1007/s41365-019-0716-x)
- [14] B. Wang, D. Han, Y. Wang *et al.*, The CEE-eTOF wall constructed with new sealed MRPC. *JINST* **15**, C08022 (2020). doi: [10.1088/1748-0221/15/08/C08022](https://doi.org/10.1088/1748-0221/15/08/C08022)
- [15] L.-K. Liu, H. Pei, Y.P. Wang *et al.*, Event plane determination from zero-degree calorimeter at the cooling-storage-ring external-target experiment. *Nucl. Sci. Tech.* **34**, 100 (2023). doi: [10.1007/s41365-023-01262-8](https://doi.org/10.1007/s41365-023-01262-8)
- [16] R. Brun, F. Bruyant, F. Carminati *et al.*, GEANT detector description and simulation tool. (1994). doi: [10.17181/CERN.MUHF.DMJ1](https://doi.org/10.17181/CERN.MUHF.DMJ1)
- [17] T. Akiba, S. Sano, T. Yanase *et al.*, Optuna: A Next-generation Hyperparameter Optimization Framework. [arXiv: 1907.10902](https://arxiv.org/abs/1907.10902)
- [18] D.M.W. Powers, Optuna: A Next-generation Hyperparameter Optimization Framework. [arXiv: 2010.16061](https://arxiv.org/abs/2010.16061)